



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### **Genomic selection for growth traits in Pacific oyster (*Crassostrea gigas*): Potential of low-density marker panels for breeding value prediction**

**Citation for published version:**

Gutierrez Silva, A, Matika, O, Bean, TP & Houston, R 2018, 'Genomic selection for growth traits in Pacific oyster (*Crassostrea gigas*): Potential of low-density marker panels for breeding value prediction', *Frontiers in genetics*, vol. 9, no. 9, 391. <https://doi.org/10.3389/fgene.2018.00391>

**Digital Object Identifier (DOI):**

[10.3389/fgene.2018.00391](https://doi.org/10.3389/fgene.2018.00391)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Frontiers in genetics

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.





# Genomic Selection for Growth Traits in Pacific Oyster (*Crassostrea gigas*): Potential of Low-Density Marker Panels for Breeding Value Prediction

Alejandro P. Gutierrez<sup>1</sup>, Oswald Matika<sup>1</sup>, Tim P. Bean<sup>2</sup> and Ross D. Houston<sup>1\*</sup>

<sup>1</sup> The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Edinburgh, United Kingdom,

<sup>2</sup> Weymouth Laboratory, Centre for Environment, Fisheries and Aquaculture Science (CEFAS), Weymouth, United Kingdom

## OPEN ACCESS

### Edited by:

Peng Xu,  
Xiamen University, China

### Reviewed by:

Xiaotong Wang,  
Ludong University, China  
Pierre Boudry,  
Institut Français de Recherche pour  
l'Exploitation de la Mer (IFREMER),  
France  
Yniv Palti,  
Cool and Cold Water Aquaculture  
Research (USDA-ARS), United States

### \*Correspondence:

Ross D. Houston  
ross.houston@roslin.ed.ac.uk

### Specialty section:

This article was submitted to  
Livestock Genomics,  
a section of the journal  
Frontiers in Genetics

Received: 08 June 2018

Accepted: 29 August 2018

Published: 19 September 2018

### Citation:

Gutierrez AP, Matika O, Bean TP and  
Houston RD (2018) Genomic  
Selection for Growth Traits in Pacific  
Oyster (*Crassostrea gigas*): Potential  
of Low-Density Marker Panels  
for Breeding Value Prediction.  
Front. Genet. 9:391.  
doi: 10.3389/fgene.2018.00391

Pacific oysters are a key aquaculture species globally, and genetic improvement via selective breeding is a major target. Genomic selection has the potential to expedite genetic gain for key target traits of a breeding program, but has not yet been evaluated in oyster. The recent development of SNP arrays for Pacific oyster (*Crassostrea gigas*) raises the opportunity to test genomic selection strategies for polygenic traits. In this study, a population of 820 oysters (comprising 23 full-sibling families) were genotyped using a medium density SNP array (23 K informative SNPs), and the genetic architecture of growth-related traits [shell height (SH), shell length (SL), and wet weight (WW)] was evaluated. Heritability was estimated to be moderate for the three traits ( $0.26 \pm 0.06$  for SH,  $0.23 \pm 0.06$  for SL and  $0.35 \pm 0.05$  for WW), and results of a GWAS indicated that the underlying genetic architecture was polygenic. Genomic prediction approaches were used to estimate breeding values for growth, and compared to pedigree based approaches. The accuracy of the genomic prediction models (GBLUP) outperformed the traditional pedigree approach (PBLUP) by ~25% for SL and WW, and ~30% for SH. Further, reduction in SNP marker density had little impact on prediction accuracy, even when density was reduced to a few hundred SNPs. These results suggest that the use of genomic selection in oyster breeding could offer benefits for the selection of breeding candidates to improve complex economic traits at relatively modest cost.

**Keywords:** genomic selection, Pacific oyster, growth, GBLUP, SNP array

## INTRODUCTION

Pacific oyster (*Crassostrea gigas*) is the most cultivated oyster species worldwide and has been introduced to many countries for aquaculture production (Troost, 2010). Global production of this species reached ~0.6 M tones in 2016 (FAO, 2018). Given its importance, several selective breeding programs based on family and mass selection have been conducted for the improvement of economically important traits such as body weight, growth rate, survival and yield (Langdon et al., 2003; Evans and Langdon, 2006; Li et al., 2011; de Melo et al., 2016), showing an improvement of the target traits after 1–5 generations. As with other aquaculture species, the recent development of genomic tools opens up the possibility for incorporating genetic markers into breeding programs via genomic selection, resulting in improved selection accuracy and genetic gain (Goddard and Hayes, 2007).

In recent years, substantial effort has been put toward the development of genomic resources for Pacific oyster, which include a reference genome assembly (Zhang et al., 2012), genetic marker databases including microsatellites (Li et al., 2003; Sekino et al., 2003) and SNPs (Sauvage et al., 2007; Fleury et al., 2009; Wang J. et al., 2015) and low density linkage maps, containing both microsatellites and SNPs (Hubert and Hedgecock, 2004; Hedgecock et al., 2015). The recent development of two oyster SNP arrays; a combined-species medium density array for Pacific oyster and European flat oyster (*O. edulis*) (Gutierrez et al., 2017) and a high density array for Pacific oyster (Qi et al., 2017); raises the opportunity of rapidly collecting genotype data for many 1000s of SNP markers dispersed throughout the genome. Moreover, a high density linkage map containing ~20 K SNPs has recently been created and aligned with the physical reference genome assembly (Gutierrez et al., 2018).

A limited number of quantitative trait locus (QTL) mapping studies have been performed to examine the genetic basis of growth related traits in Pacific oyster (Hedgecock et al., 2007; Guo et al., 2012; Wang and Li, 2017), generally indicating that these traits are polygenic in nature. With the recently developed genotyping resources, GWAS might have the potential to address some of the drawbacks that QTL mapping (based on linkage) has, particularly related to marker density. Moreover, GWAS are based on a population level linkage disequilibrium between markers and QTL, which could potentially facilitate the application of marker-assisted selection (MAS) in breeding programs. However, MAS based on polygenic traits is likely to be ineffective due to capturing only a small proportion of the genetic variation in the trait. Therefore, genomic selection may be a promising avenue for incorporating markers into shellfish breeding. In genomic selection, genome-wide SNP markers are used to generate a genomic relationship matrix which is utilized to predict genomic estimated breeding values (GEBVs) for individuals without phenotypes, based on training of the genomic prediction equation in a reference population with both phenotypes and genotypes (Goddard and Hayes, 2007). Given that genomic selection can be used to accurately predict breeding values even in the absence of trait or pedigree information, it may have high potential for oysters where routine pedigree recording can be difficult for two reasons. The first is that maintaining physically separate families is logistically difficult and expensive, and tagging of juvenile oysters is challenging. The second is that genotyping errors for traditional marker assays such as microsatellites are common, and several examples of incorrect pedigree assignment have been described, thought to be due to the high frequency of null alleles, ranging from 16 to 51% (Launey and Hedgecock, 2001; Hedgecock et al., 2004; Reece et al., 2004). Encouraging, the use of SNPs from the oyster SNP array platform was shown to be successful in the parental assignment in a limited number of Pacific oyster families (Gutierrez et al., 2018). The advantages of genomic selection over traditional pedigree-based approaches in terms of accuracy of the predictions observed for both livestock and aquaculture species for polygenic traits have been described for several aquaculture species (Ødegård et al., 2014; Tsai et al., 2015; Dou et al., 2016; Palaikostas et al., 2016; Correa et al., 2017; Vallejo et al., 2017a). However, despite the

importance of Pacific oyster to global aquaculture, no studies have yet evaluated the potential of genomic selection for breeding value prediction in this species.

The primary aim of the current study was to evaluate the potential use of genomic prediction in a population of Pacific oysters derived from a commercial hatchery. Several growth-related traits were evaluated as exemplar polygenic traits, including shell length (SL), shell height (SH), and wet weight (WW). The impact of SNP marker density on genomic prediction accuracy was evaluated, alongside strategies for selecting low density panels for potential improvement of genotyping cost-efficiency.

## MATERIALS AND METHODS

### Source of Oysters

The population used in this study were derived from crosses between broodstock from a commercial oyster hatchery (Guernsey Sea Farms, United Kingdom) and were a subset of the samples used for analysis of resistance to *Ostreid Herpesvirus*, as described in Gutierrez et al. (2018). There were two sets of oyster crosses used in the study. The first comprised three pair crosses that were created at Cefas (from 3 sires and 2 dams) and then reared in separate tanks. Larvae were held in 5 L tanks with daily water renewal, daily feeding and constant aeration. Post settling (at roughly 2 weeks) these were moved to 10 L tanks with a constant flow of water and feed. Spat were handled every 2–3 weeks, when the tanks were cleaned. Feed was provided at the rate recommend in the manual for hatchery culture of bivalve mollusc, according to the spat density and water volume of each tank. Larvae were fed a mix of *Chaetoceros*, T-Isochrysis and Pavlova algae. The remaining crosses (from 14 sires and 14 dams) were obtained as spat from a mass spawning at Guernsey Sea Farms (GSF). Prior to settling, larvae were held in upwelling 2 L bottle system, with aeration and constant supply of feed according to standard hatchery procedure. Post settling, oysters were held at GSF for 3 weeks then delivered overnight to Cefas where they were held on mesh upwelling system in a large 60 L tank, with a constant flow of water.

For both groups of oysters reared at Cefas (from larval stage for the first group, from post-settlement stage for the second group), spat were handled every 2–3 weeks, when the tanks were cleaned. The oysters were drip fed a constant supply of mixed food including Pavlova, T-Isochrysis and Tetraselmis. Throughout the experiment, all sea water was filtered, UV treated, mixed to a salinity of 25 ppt with RO water, and aerated prior to use. Feed was provided at the rate recommend in the manual for hatchery culture of bivalve mollusc, according to the density of each tank. On several occasions, where algal stocks were low, food was supplemented with shellfish diet 1,800<sup>1</sup>. Measurements were taken at approximately 6 months. The differences in early life environment may have affected growth rate and were therefore accounted for in the statistical model described below. Parental assignment was performed as described

<sup>1</sup><http://reedmariculture.com/>

in Gutierrez et al. (2018), resulting in the identification of 23 different full-sibling families in the population. All animals were reared in accordance with the United Kingdom Home Office regulations regarding the use of animals in experiments. The trial was carried out at the Centre for Environment, Fisheries and Aquaculture Science (Cefas, United Kingdom).

## Phenotypic Measurements

Shell measurements (SL and SH) were taken following a standard protocol for the measurement of oyster shells (Galtsoff, 1964). Wet weight (WW) was recorded during the DNA extraction procedures where the whole animal (excluding shell) was used.

## SNP Array Genotyping

Genome-wide SNP data were generated using the recently developed Affymetrix SNP array for oysters (Gutierrez et al., 2017), as described in Gutierrez et al. (2018). Briefly, genomic DNA was extracted from the whole oyster (minus the shell) using the RealPure genomic DNA extraction kit (Valencia, Spain), quantified on a Qubit fluorometer (Invitrogen) and the DNA integrity was checked on a 1% agarose gel. Array genotyping was carried out at Edinburgh Genomics, and quality control was performed using the Axiom Analysis Suite v2.0.0.35, following the “best practices workflow” with a sample and SNP call threshold of 90%. These settings resulted in 23,388 SNPs classified as good quality and therefore retained for downstream analyses. Post-filtering, the final dataset comprised 820 individuals with genotype and phenotype data.

## Genetic Parameter Estimation

Genetic parameters for the resistance traits were estimated using a linear mixed model approach fitting animal as a random effect using ASReml 4 (Gilmour et al., 2015) with the following model:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (1)$$

where  $\mathbf{y}$  is the observed trait,  $\mathbf{u}$  is the vector of additive genetic effects,  $\mathbf{b}$  is the vector of fixed effect of tank,  $\mathbf{e}$  is the residual error, and  $\mathbf{X}$  and  $\mathbf{Z}$  the corresponding incidence matrices for fixed effects and additive genetic effects, respectively. The (co)variance structure for the genetic effect was calculated either using the pedigree matrix ( $\mathbf{A}$ ) (i.e.,  $\mathbf{u} \sim N(0, \mathbf{A}\sigma_a^2)$ ) or genomic matrix ( $\mathbf{G}$ )  $N(0, \mathbf{G}\sigma_a^2)$ , where  $\sigma^2$  is the genetic variance. Hence, the narrow sense heritability was estimated by the additive genetic variance and total phenotypic variance, as follows:

$$h^2 = \sigma_a^2 / \sigma_p^2 \quad (2)$$

where  $\sigma_a^2$  is the additive genetic variance and  $\sigma_p^2$  is the total phenotypic variance which is a sum of  $\sigma_a^2 + \sigma_e^2$ . The genomic relationship matrix used for the analysis was obtained according to VanRaden (2008) using the GenABEL package (Aulchenko et al., 2007) and inverted using a standard “R” function. The fixed effect of “tank” partially accounted for the differences in early life conditions between the pair-cross and batch-spawned oysters, which were held in separate tanks at Cefas during post-settlement rearing.

## Genome-Wide Association Studies

The GWAS were performed for the three growth-related traits using two approaches, first using the GenABEL package (Aulchenko et al., 2007) in R and also genomic BLUP analysis implemented in BLUPF90 software (Misztal et al., 2002). The genotype data was filtered as part of quality control by using the *check.markers* module to exclude SNPs with a minor allele frequency (MAF)  $<0.05$ , call rate  $<0.90$  and significantly deviated from Hardy–Weinberg Equilibrium  $<1 \times 10^{-6}$ , leaving 13,278 SNPs for downstream analyses. Association analyses were performed using the family-based score test for association (FASTA) using the *mmscore* function (Chen and Abecasis, 2007) with the mixed linear model (MLM) approach used to avoid potential false positive associations due to population structure. Genotype data were used to calculate the genomic kinship matrix which was fitted in the model alongside the top four principal components as covariates to account for population structure. The genome-wide significance threshold was set to  $3.76 \times 10^{-6}$  as determined by Bonferroni correction ( $0.05/N$ ), where  $N$  represents the number of QC-filtered SNPs across the genome, while the suggestive threshold was set as  $3.76 \times 10^{-5}$  ( $0.5/N$ ), i.e., allowing 0.5 false positive per genome scan. For the BLUPF90 approach, the same data previously filtered by GenABEL were used. Model (1) was fitted using the genomic ( $\mathbf{G}$ ) relationship matrix that was created according to VanRaden (2008). In this case, windows of 10 adjacent (not overlapping) SNPs based on the linkage map position were created using POSTGSF90 (Aguilar et al., 2014). It has been shown that the use of a higher number of SNPs (as SNP windows) in the analysis of quantitative traits could capture the QTL effect more accurately than a single SNP (Habier et al., 2011). Even though recent studies have argue that the use of a higher number of markers in the window should provide better power for the detection of QTL (Wang H. et al., 2014; Gonzalez-Pena et al., 2016; Vallejo et al., 2017b). We chose to only use 10 consecutive SNPs (as windows) to reduce the possibilities of wrong SNP position, given that numerous assembly errors have been described within the oyster genome (Hedgecock et al., 2015; Gutierrez et al., 2018), and therefore the SNP order may not be fully accurate.

## Genomic Prediction

For the estimation of genomic prediction values, the genotype data was filtered to allow markers with a (MAF)  $>0.01$ , which resulted in a higher number of markers in the analysis (16,079 SNPs). Estimated breeding values were obtained using either pedigree-based BLUP (PBLUP) or Genomic best linear unbiased prediction (GBLUP) using the linear model described above. The accuracy of genomic selection was estimated by fivefold cross validation (training set 80%, validation set 20%), which were repeated five times. Phenotypes from the validation population were masked and breeding values were estimated using ASReml 4 using the linear mixed model described above (1). Prediction accuracy was calculated as the correlation between the predicted EBVs of the validation set and the actual phenotypes divided by the square root of the heritability estimated in the validation population  $[\sim r(y_1, y_2)/h]$ . Mean prediction accuracy values



obtained from the different sets were computed and compared between the pedigree and genomic approaches.

Two different strategies for evaluating the potential of lower marker densities for genomic prediction were applied. First, the low density SNP panel for use in the computing the genomic relationship matrix was selected by a progressive increase of the MAF threshold from 0.01 to 0.475 resulting in a progressive reduction in number of markers (as shown in **Table 2**); Secondly, the low density SNP panel was selected using a strategy of random “thinning” of SNPs from the full dataset (15, 10, 5, 2.5, 1, and down to 100 SNPs).

## Data Availability

Genotype data corresponding to these samples has already been made publicly available as **Supplementary Material** by Gutierrez et al. (2018). The combined-species medium density SNP array for oysters can be ordered from Thermo Fisher Scientific (Waltham, CA, United States).

## RESULTS AND DISCUSSION

### Trait Summary and Heritability

The mean and standard deviation values for the growth-related traits were  $11.27 \pm 2.9$ ,  $34.94 \pm 28.33$ , and  $8.3 \pm 1.97$  for SH, WW, and SL, respectively. The genetic correlation between the traits was high  $>0.9$  for all traits and the phenotypic correlation slightly lower (**Table 1**). Moderate heritability values were estimated for all traits, ranging from 0.20 (pedigree-based) for length, to 0.35 for weight (G-matrix based), as shown in **Table 1**. Previous studies have described significant heritability values for growth traits in Pacific oyster, albeit these estimates range in magnitude. For instance, Dégremont et al. (2007) reported the heritability estimates for weight that ranged from 0.07 to 0.15 in 6–8 months old *C. gigas*, while Kong et al. (2015) and Lannan (1972) reported values for weight at  $\sim 0.35$  in older samples of the same species. Other studies have reported heritability estimates for weight ranging from 0.1 to 0.5 in *C. gigas* (Hedgecock et al., 1991; Sheridan, 1997; Langdon et al., 2003; Evans and Langdon, 2006). Our results are consistent with heritability estimates described for shell measurements, e.g., Li et al. (2011) reported values of 0.149–0.402 for SH in *C. gigas* at 12 months of age, while Kong et al. (2015) described estimates of 0.49 for SH and 0.36 for SL, and Xu et al. (2017) reported heritability values of 0.18 for SH and 0.25 for SL. These results imply that growth can potentially be improved by selective breeding. Most oyster breeding programs are focused on increasing flesh weight, but since a high genetic correlation between the three analyzed traits ( $>0.90$ ) was observed, selection (either pedigree or genomic based) for any one of the traits is likely to co-select for improvement of the other traits. High genetic and phenotypic correlations between these traits have also been observed previously (Kong et al., 2015), and (combinations of) these traits can be used to predict soft-body WW which is a main breeding objective for oysters (Wang X. et al., 2014).

It should be noted that the oysters used to assess early-life growth in the current were potentially influenced by differences

in early life environment, and by the fact that they were survivors of an OsHV-1 challenge experiment (Gutierrez et al., 2018). However, there was no evidence of genetic correlation between the traits of survival and the growth-related traits (data not shown), which is consistent with the findings of Dégremont et al. (2007). The early life conditions of the batch-spawned oysters and the laboratory spawned oysters were different, and this was partially accounted for by the inclusion of the fixed effect of “tank” in the statistical model (since the batch-spawned and pair-cross oysters were held in separate tanks at Cefas post-settlement). While this fixed effect was not significant, these early life differences may have impacted on the results, since genotype by environment interaction has been shown to be important for growth traits in Pacific oyster (Langdon et al., 2003; Dégremont et al., 2007). In addition, the disease challenge experiment may have had a minor influence on size, since growth during challenge may not be the same trait as growth in the absence of disease challenge. While the traits measured may not be a reliable indicator of oyster growth under commercially relevant conditions, they were heritable and polygenic and therefore served purpose for testing genomic prediction approaches.

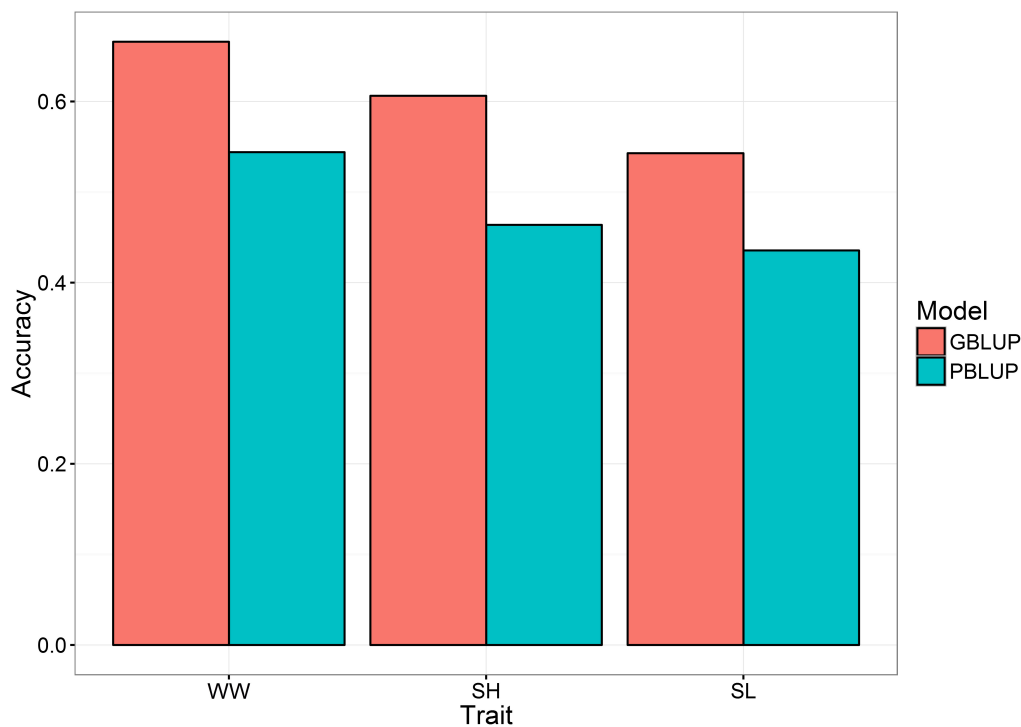
### Genome-Wide Association Studies (GWAS)

There was no significant or suggestive association detected by GenABEL between any SNP and any of the three analyzed traits (**Supplementary Table S1** and **Supplementary Figure S1**). Additionally, the proportion of variance explained (PVE) by each SNP was low ( $\sim 2\%$ ) highlighting the absence of any major QTL controlling growth traits in this population. Moreover, BLUPF90 analysis based on consecutive (not overlapping) SNP windows did not detect major QTL in any of the three traits, although a suggestive QTL was detected for SL in a window that explained 1.48% of the genetic variance and located on LG 1 (48.61 cM) (**Supplementary Table S2** and **Supplementary Figure S2**), Worth noting that 2 windows located at LG 7 (73.23 cM) and LG 6 (39.94–40.82 cM) were found among the 10 highest scores for the three analyzed traits. These results suggest a polygenic nature of growth-related traits in Pacific oyster, controlled by

**TABLE 1 |** Genetic parameter estimates for the growth-related traits in the Pacific oyster samples.

	SH	SL	WW
Mean (s.d)	11.27 (2.9)	8.3 (1.9)	34.94 (28.3)
<b>HERITABILITY<sup>b</sup></b>			
G-matrix	0.26 (0.05)	0.23 (0.06)	0.35 (0.05)
A-Matrix	0.23 (0.12)	0.20 (0.11)	0.31 (0.13)
<b>CORRELATION<sup>a</sup></b>			
SH	–	0.83 (0.01)	0.78 (0.02)
SL	0.95 (0.04)	–	0.74 (0.02)
WW	0.92 (0.06)	0.90 (0.06)	–

<sup>a</sup>Genetic correlation was estimated based on the genomic relationship matrix and values are shown below the diagonal, while phenotypic correlation values are shown above the diagonal. <sup>b</sup>Heritability was estimated based on the genomic relationship matrix (G-matrix) and the pedigree.



**FIGURE 1** | Mean accuracies for GBLUP and PBLUP for the three analyzed traits. WW, wet weight; SH, shell height; SL, shell length.

many loci with small effects as it has been observed in many aquaculture species such as Atlantic salmon (Gutierrez et al., 2015; Tsai et al., 2015; Yoshida et al., 2017), rainbow trout (Wringe et al., 2010), common carp (Palaikostas et al., 2018), grouper (Yu et al., 2018), turbot (Sánchez-Molano et al., 2011), Asian seabass (Wang L. et al., 2015), and also oysters (Guo et al., 2012; Li and He, 2014).

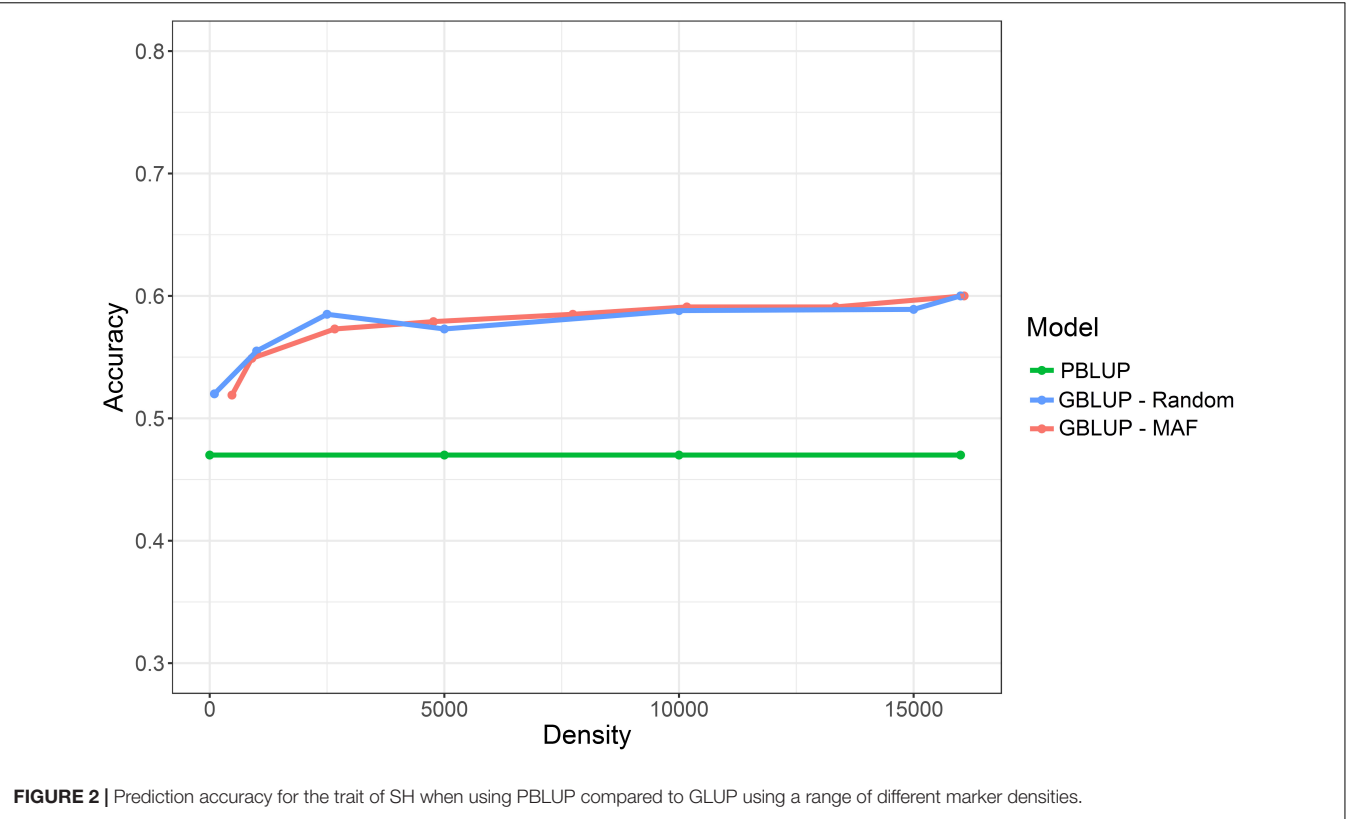
To inform a strategy for deployment of genetic markers in commercial aquaculture breeding programs, it is important to first define the genetic architecture of the target trait(s) in question. Marker assisted selection (MAS) is most appropriate when major QTL explain a high proportion of genetic variation in the trait, enabling use of a small panel of QTL-linked markers to supplement pedigree-based selection. A good example of this is resistance to the Infectious Pancreatic Necrosis virus (IPNV) in Atlantic salmon, where most genetic variation is explained by a single QTL (Houston et al., 2008; Moen et al., 2009). However, most growth-related traits (and other traits of economic importance to aquaculture production) are polygenic, and therefore genomic selection incorporating all markers to predict breeding values is likely to be a more effective approach.

## Genomic Prediction

Genomic prediction accuracy for the three traits were tested using genotype information from 16,076 markers that passed the QC filter ( $MAF > 0.01$ ,  $CallRate > 0.9$ ). Animals were randomly split into training (80%) and validation (20%) sets for cross-validation, and this was repeated five times. The genomic prediction accuracy results highlighted that the higher prediction accuracies

were obtained using the genomic information (G-matrix) than using the pedigree information (A-matrix) (**Figure 1**), with an increase of ~25% for WW (from 0.54 PBLUP to 0.67 GBLUP) and SL (from 0.44 PBLUP to 0.54 GBLUP) to ~30% for SH (from 0.47 PBLUP to 0.60 GBLUP). These results are in agreement with published literature which shows the increase in breeding value prediction accuracy using genomic vs. pedigree prediction in aquaculture species, e.g., Atlantic salmon (Ødegård et al., 2014; Yoshida et al., 2017; Barría et al., 2018; Robledo et al., 2018), rainbow trout (Vallejo et al., 2017a; Yoshida et al., 2018), sea bream (Palaikostas et al., 2016). For growth traits in particular, studies in Atlantic salmon (Tsai et al., 2015), common carp (Palaikostas et al., 2018), and Pacific white shrimp (Wang et al., 2017) have reported an increase in the prediction accuracies by the use of genomic information. The results of the current study highlight the potential of genomic selection for economic traits in oysters, albeit the cost may be prohibitively expensive and more cost-effective genotyping strategies may be required for feasibility of commercial application.

To evaluate the effect of marker density on genomic prediction accuracy, two strategies of obtaining lower density SNP panels were applied. The first used progressive increase of minor allele frequency (MAF) threshold, resulting in progressive decrease in SNP number (as described in Robledo et al. (2018)). The second involved choosing subsets of SNPs for the low density panels at random. Using both approaches, significant reduction in SNP density had little impact on prediction accuracy until marker densities dropped below ~2,500 SNPs. With the MAF approach, the genomic prediction accuracies obtained using the



lower density SNP panels ranged from 0.59 (MAF > 0.05; 13,337 SNPs) to 0.52 (MAF > 0.475; 474 SNPs) (Table 2 and Figure 2). Using the random subsets the prediction accuracies ranged from 0.59 (15,000 SNPs) to 0.52 (100 SNPs). In all cases, the genomic prediction significantly outperformed pedigree prediction, even for low SNP densities, which has positive implications for future use of low-cost low-density SNP panels for genomic selection. The oyster genome is of a moderate size ~0.6 gb (Zhang et al., 2012) and that could be related to the low number of markers needed to high prediction accuracies. Additionally, it is important to note that since this a relatively small sample set that contains a relatively limited number of families of known structure, a high level of relatedness between the training and validation sets is expected. A total of 23 nuclear families were included in the

study, for which 15 dams and 16 sires were effective breeders, and resulted in a mix of full sibling and half sibling families (see Supplementary File S1). This may result in high levels of linkage disequilibrium across large chromosome segments, which could influence the genomic prediction accuracy estimates at low SNP densities. Nonetheless, this may be representative of typical aquaculture breeding schemes that utilize large full-sib families for “sib testing,” and where genomic prediction estimates with low density markers can give high prediction accuracy (Lillehammer et al., 2013; Tsai et al., 2015). Current limitations of the use of genomic selection in aquaculture companies relate to the cost of phenotype recording and genotyping, and the latter could be eased by the use of low cost genotyping which is directly related to number of SNPs to be typed. Our results suggest that

**TABLE 2 |** Genomic prediction values obtained for SH using decreasing marker densities.

Method	Approach	SNP N	Accuracy	Approach	SNP N	Accuracy
PBLUP	Pedigree	–	0.47			
GBLUP	MAF 0.01	16,076	0.6	Random	16,076	0.6
GBLUP	MAF 0.05	13,337	0.59	Random	15,000	0.59
GBLUP	MAF 0.1	10,167	0.59	Random	10,000	0.59
GBLUP	MAF 0.15	7,738	0.59	Random	5,000	0.57
GBLUP	MAF 0.25	4,768	0.58	Random	2,500	0.59
GBLUP	MAF 0.35	2,664	0.57	Random	1,000	0.56
GBLUP	MAF 0.45	898	0.55	Random	100	0.52
GBLUP	MAF 0.475	474	0.52			

low density SNP panels, as few as several hundred SNPs, may be sufficient to achieve the asymptote of prediction accuracy in a mixed family population of oysters that may be typical of an oyster breeding population.

## CONCLUSION

A recently developed medium density SNP array was used to evaluate the efficacy of different strategies for genomic prediction in a population of oysters derived from a commercial hatchery. Three growth-related traits were analyzed as exemplar polygenic traits (which was confirmed by GWAS). The three traits were found to be moderately heritable and showed high genetic correlation. Prediction accuracy for all traits was substantially higher using genomic prediction than pedigree-based prediction. Reduction in SNP marker density had little impact on prediction accuracy when the lower density SNP panels were chosen at random, implying only a fraction of the SNPs are required to obtain a marked increase in accuracy relative to pedigree-based prediction. These results suggest that the use of cost-effective genomic selection in oyster breeding could bring major benefits for the selection of polygenic traits, and may have commercial value for traits which are difficult to measure e.g., disease resistance.

## REFERENCES

- Aguilar, I., Misztal, I., Tsuruta, S., Legarra, A., and Wang, H. (2014). "PREGSF90 – POSTGSF90: computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs," in *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production*, Vancouver, BC. doi: 10.1093/bioinformatics/btm108
- Aulchenko, Y. S., Ripke, S., Isaacs, A., and Van Duijn, C. M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23, 1294–1296. doi: 10.1534/g3.118.200053
- Barria, A., Christensen, K. A., Yoshida, G. M., Correa, K., Jedlicki, A., Lhorente, J. P., et al. (2018). Genomic predictions and genome-wide association study of resistance against *Piscirickettsia salmonis* in coho salmon (*Oncorhynchus kisutch*) using ddRAD sequencing. *G3* 8, 1183–1194. doi: 10.1534/g3.118.200053
- Chen, W.-M., and Abecasis, G. R. (2007). Family-based association tests for genomewide association scans. *Am. J. Hum. Genet.* 81, 913–926. doi: 10.1186/s12711-017-0291-8
- Correa, K., Banger, R., Figueroa, R., Lhorente, J. P., and Yáñez, J. M. (2017). The use of genomic information increases the accuracy of breeding value predictions for sea louse (*Caligus rogercresseyi*) resistance in Atlantic salmon (*Salmo salar*). *Genet. Sel. Evol.* 49:15. doi: 10.1186/s12711-017-0291-8
- de Melo, C. M. R., Durland, E., and Langdon, C. (2016). Improvements in desirable traits of the Pacific oyster, *Crassostrea gigas*, as a result of five generations of selection on the West Coast, USA. *Aquaculture* 460, 105–115. doi: 10.1016/j.aquaculture.2006.10.025
- Dégremont, L., Ernande, B., Bédier, E., and Boudry, P. (2007). Summer mortality of hatchery-produced Pacific oyster spat (*Crassostrea gigas*). I. Estimation of genetic parameters for survival and growth. *Aquaculture* 262, 41–53. doi: 10.1038/srep19244
- Dou, J., Li, X., Fu, Q., Jiao, W., Li, Y., Li, T., et al. (2016). Evaluation of the 2b-RAD method for genomic selection in scallop breeding. *Sci. Rep.* 6:19244. doi: 10.1038/srep19244
- Evans, S., and Langdon, C. (2006). Direct and indirect responses to selection on individual body weight in the Pacific oyster (*Crassostrea gigas*). *Aquaculture* 261, 546–555.
- FAO (2018). *Food and Agriculture Organization Statistical Yearbook*. Rome: FAO. doi: 10.1186/1471-2164-10-341
- Fleury, E., Huvet, A., Lelong, C., De Lorgeril, J., Boulo, V., Gueguen, Y., et al. (2009). Generation and analysis of a 29,745 unique Expressed Sequence Tags from the Pacific oyster (*Crassostrea gigas*) assembled into a publicly accessible database: the GigasDatabase. *BMC Genomics* 10:341. doi: 10.1186/1471-2164-10-341
- Galtsoff, P. S. (1964). The American oyster *Crassostrea virginica* (Gmelin). *US Fish Wildl. Serv. Fish. Bull.* 64, 1–480. doi: 10.1111/j.1439-0388.2007.00702.x
- Gilmour A., Gogel, B., Cullis, B., Welham S., and Thompson R. (2015). *ASReml User Guide Release 4.1 Structural Specification*. Hemel Hempstead: VSN International Ltd.
- Goddard, M. E., and Hayes, B. J. (2007). Genomic selection. *J. Anim. Breed. Genet.* 124, 323–330. doi: 10.3389/fgene.2016.00203
- Gonzalez-Pena, D., Gao, G., Baranski, M., Moen, T., Cleveland, B. M., Kenney, P. B., et al. (2016). Genome-wide association study for identifying loci that affect fillet yield, carcass, and body weight traits in rainbow trout (*Oncorhynchus mykiss*). *Front. Genet.* 7:203. doi: 10.3389/fgene.2016.00203
- Guo, X., Li, Q., Wang, Q. Z., and Kong, L. F. (2012). Genetic mapping and QTL analysis of growth-related traits in the Pacific oyster. *Mar. Biotechnol.* 14, 218–226. doi: 10.1007/s10126-011-9405-4
- Gutierrez, A. P., Bean, T. P., Hooper, C., Stenton, C. A., Sanders, M. B., Paley, R. K., et al. (2018). A genome-wide association study for host resistance to ostreid herpesvirus in Pacific oysters (*Crassostrea gigas*). *G3* 8, 1273–1280. doi: 10.1534/g3.118.200113
- Gutierrez, A. P., Turner, F., Gharbi, K., Talbot, R., Lowe, N. R., Peñaloza, C., et al. (2017). Development of a medium density combined-species SNP array for Pacific and European oysters (*Crassostrea gigas* and *Ostrea edulis*). *G3* 7, 2209–2218. doi: 10.1534/g3.117.041780
- Gutierrez, A. P., Yáñez, J. M., Fukui, S., Swift, B., and Davidson, W. S. (2015). Genome-wide association study (GWAS) for growth rate and age at

## AUTHOR CONTRIBUTIONS

AG and RH conceived the study. TB and RH designed the experimental structure. TB established and performed the experimental challenge. AG performed DNA extractions, genotype processing, and parentage assignment. AG and OM performed the quantitative genetic analyses. All authors contributed to drafting the manuscript.

## FUNDING

The authors gratefully acknowledge funding from BBSRC and NERC under the United Kingdom Aquaculture Initiative (BB/M026140/1 and NE/P010695/1) in addition to BBSRC Institute Strategic Program Grants (BBS/E/D/20002172 and BBS/E/D/30002275). Edinburgh Genomics was partly supported through core grants from NERC (R8/H10/56), MRC (MR/K001744/1), and BBSRC (BB/J004243/1).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00391/full#supplementary-material>



- sexual maturation in Atlantic salmon (*Salmo salar*). *PLoS One* 10:e0119730. doi: 10.1371/journal.pone.0119730
- Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186. doi: 10.1186/1471-2105-12-186
- Hedgecock, D., Cooper, K., and Hershberger, W. (1991). Genetic and environmental components of variance in harvest body size among pedigreed Pacific oysters *Crassostrea gigas* from controlled crosses. *J. Shellfish Res.* 10, 516.
- Hedgecock, D., Li, G., Hubert, S., Bucklin, K., and Ribes, V. (2004). Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the Pacific oyster, *Crassostrea gigas*. *J. Shellfish Res.* 23, 379–386. doi: 10.1016/j.aquaculture.2007.07.085
- Hedgecock, D., Perry, G. M. L., and Voigt, M. L. (2007). Mapping heterosis QTL in the Pacific oyster *Crassostrea gigas*. *Aquaculture* 272, S267–S268. doi: 10.1534/g3.115.019570
- Hedgecock, D., Shin, G., Gracey, A. Y., Den Berg, D. V., and Samanta, M. P. (2015). Second-generation linkage maps for the Pacific oyster *Crassostrea gigas* reveal errors in assembly of genome scaffolds. *G3* 5, 2007–2019. doi: 10.1534/g3.115.019570
- Houston, R. D., Haley, C. S., Hamilton, A., Guy, D. R., Tinch, A. E., Taggart, J. B., et al. (2008). Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). *Genetics* 178, 1109–1115. doi: 10.1534/genetics.104.027342
- Hubert, S., and Hedgecock, D. (2004). Linkage maps of microsatellite DNA markers for the Pacific oyster *Crassostrea gigas*. *Genetics* 168, 351–362. doi: 10.1111/are.12205
- Kong, N., Li, Q., Yu, H., and Kong, L. F. (2015). Heritability estimates for growth-related traits in the Pacific oyster (*Crassostrea gigas*) using a molecular pedigree. *Aquac. Res.* 46, 499–508. doi: 10.1016/S0044-8486(02)00621-X
- Langdon, C., Evans, F., Jacobson, D., and Blouin, M. (2003). Yields of cultured Pacific oysters *Crassostrea gigas* Thunberg improved after one generation of selection. *Aquaculture* 220, 227–244.
- Lannan, J. E. (1972). Estimating heritability and predicting response to selection for the Pacific oyster, *Crassostrea gigas*. *Proc. Natl. Shellfish Assoc.* 62, 62–66.
- Launey, S., and Hedgecock, D. (2001). High genetic load in the Pacific oyster *Crassostrea gigas*. *Genetics* 159, 255–265. doi: 10.1046/j.1471-8286.2003.00406.x
- Li, G., Hubert, S., Bucklin, K., Ribes, V., and Hedgecock, D. (2003). Characterization of 79 microsatellite DNA markers in the Pacific oyster *Crassostrea gigas*. *Mol. Ecol. Notes* 3, 228–232. doi: 10.1007/s12562-011-0369-0
- Li, Q., Wang, Q., Liu, S., and Kong, L. (2011). Selection response and realized heritability for growth in three stocks of the Pacific oyster *Crassostrea gigas*. *Fish. Sci.* 77, 643–648. doi: 10.1371/journal.pone.0111707
- Li, Y., and He, M. (2014). Genetic mapping and QTL analysis of growth-related traits in *Pinctada fucata* using restriction-site associated DNA sequencing. *PLoS One* 9:e111707. doi: 10.1371/journal.pone.0111707
- Lillehammer, M., Meuwissen, T. H. E., and Sonesson, A. K. (2013). A low-marker density implementation of genomic selection in aquaculture using within-family genomic breeding values. *Genet. Sel. Evol.* 45:39. doi: 10.1186/1297-9686-45-39
- Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., and Lee, D. (2002). “BLUPF90 and related programs (BGF90),” in *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, Montpellier. doi: 10.1186/1471-2164-10-368
- Moen, T., Baranski, M., Sonesson, A. K., and Kjøglum, S. (2009). Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait. *BMC Genomics* 10:368. doi: 10.1186/1471-2164-10-368
- Ødegård, J., Moen, T., Santi, N., Korsvoll, S. A., Kjøglum, S., and Meuwissen, T. H. E. (2014). Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). *Front. Genet.* 5:402. doi: 10.3389/fgene.2014.00402
- Palaiokostas, C., Ferrarese, S., Franch, R., Houston, R. D., and Bargelloni, L. (2016). Genomic prediction of resistance to pasteurellosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. *G3* 6, 3693–3700. doi: 10.1534/g3.116.035220
- Palaiokostas, C., Kocour, M., Prchal, M., and Houston, R. D. (2018). Accuracy of genomic evaluations of juvenile growth rate in common carp (*Cyprinus carpio*) using genotyping by sequencing. *Front. Genet.* 9:82. doi: 10.3389/fgene.2018.00082
- Qi, H., Song, K., Li, C., Wang, W., Li, B., Li, L., et al. (2017). Construction and evaluation of a high-density SNP array for the Pacific oyster (*Crassostrea gigas*). *PLoS One* 12:e0174007. doi: 10.1371/journal.pone.0174007
- Reece, K. S., Ribeiro, W. L., Gaffney, P. M., Carnegie, R. B., and Allen, J. S. K. (2004). Microsatellite marker development and analysis in the eastern oyster (*Crassostrea virginica*): confirmation of null alleles and non-mendelian segregation ratios. *J. Hered.* 95, 346–352. doi: 10.1093/jhered/esh058
- Robledo, D., Matika, O., Hamilton, A., and Houston, R. D. (2018). Genome-wide association and genomic selection for resistance to amoebic gill disease in Atlantic salmon. *G3* 8, 1195–1203. doi: 10.1534/g3.118.200075
- Sánchez-Molano, E., Cerna, A., Toro, M. A., Bouza, C., Hermida, M., Pardo, B. G., et al. (2011). Detection of growth-related QTL in turbot (*Scophthalmus maximus*). *BMC Genomics* 12:473. doi: 10.1186/1471-2164-12-473
- Sauvage, C., Bierre, N., Lapègue, S., and Boudry, P. (2007). Single Nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. *Gene* 406, 13–22. doi: 10.1016/j.gene.2007.05.011
- Sekino, M., Hamaguchi, M., Aranishi, F., and Okoshi, K. (2003). Development of novel microsatellite DNA markers from the Pacific oyster *Crassostrea gigas*. *Mar. Biotechnol.* 5, 227–233. doi: 10.1007/s10126-002-0104-z
- Sheridan, A. K. (1997). Genetic improvement of oyster production—a critique. *Aquaculture* 153, 165–179. doi: 10.1016/S0044-8486(97)00024-0
- Troost, K. (2010). Causes and effects of a highly successful marine invasion: case-study of the introduced Pacific oyster *Crassostrea gigas* in continental NW European estuaries. *J. Sea Res.* 64, 145–165. doi: 10.1016/j.seares.2010.02.004
- Tsai, H.-Y., Hamilton, A., Tinch, A. E., Guy, D. R., Gharbi, K., Stear, M. J., et al. (2015). Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics* 16:969. doi: 10.1186/s12864-015-2117-9
- Vallejo, R. L., Leeds, T. D., Gao, G., Parsons, J. E., Martin, K. E., Evenhuis, J. P., et al. (2017a). Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. *Genet. Sel. Evol.* 49:17. doi: 10.1186/s12711-017-0293-6
- Vallejo, R. L., Liu, S., Gao, G., Fragomeni, B. O., Hernandez, A. G., Leeds, T. D., et al. (2017b). Similar genetic architecture with shared and unique quantitative trait loci for bacterial cold water disease resistance in two rainbow trout breeding populations. *Front. Genet.* 8:156. doi: 10.3389/fgene.2017.00156
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- Wang, H., Misztal, I., Aguilar, I., Legarra, A., Fernando, R. L., Vitezica, Z., et al. (2014). Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. *Front. Genet.* 5:134. doi: 10.3389/fgene.2014.00134
- Wang, J., and Li, Q. (2017). Characterization of novel EST-SNP markers and their association analysis with growth-related traits in the Pacific oyster *Crassostrea gigas*. *Aquac. Int.* 25, 1707–1719. doi: 10.1007/s10499-017-0142-1
- Wang, J., Qi, H., Li, L., Que, H., Wang, D., and Zhang, G. (2015). Discovery and validation of genic single nucleotide polymorphisms in the Pacific oyster *Crassostrea gigas*. *Mol. Ecol. Resour.* 15, 123–135. doi: 10.1111/1755-0998.12278
- Wang, L., Wan, Z. Y., Bai, B., Huang, S. Q., Chua, E., Lee, M., et al. (2015). Construction of a high-density linkage map and fine mapping of QTL for growth in Asian seabass. *Sci. Rep.* 5:16358. doi: 10.1038/srep16358
- Wang, Q., Yu, Y., Yuan, J., Zhang, X., Huang, H., Li, F., et al. (2017). Effects of marker density and population structure on the genomic prediction accuracy for growth trait in Pacific white shrimp *Litopenaeus vannamei*. *BMC Genet.* 18:45. doi: 10.1186/s12863-017-0507-5
- Wang, X., Liu, T., Liu, Y., and Feng, P. (2014). An arithmetic index based on shell height, length, and width, for potential selection of soft-body wet weight in Pacific oyster, *Crassostrea gigas*. *Isr. J. Aquac.* 66:4.
- Wringe, B. F., Devlin, R. H., Ferguson, M. M., Moghadam, H. K., Sakhrani, D., and Danzmann, R. G. (2010). Growth-related quantitative trait loci in domestic and wild rainbow trout (*Oncorhynchus mykiss*). *BMC Genet.* 11:63. doi: 10.1186/1471-2156-11-63

- Xu, L., Li, Q., Yu, H., and Kong, L. (2017). Estimates of heritability for growth and shell color traits and their genetic correlations in the black shell strain of pacific oyster *Crassostrea gigas*. *Mar. Biotechnol.* 19, 421–429. doi: 10.1007/s10126-017-9772-6
- Yoshida, G. M., Bangera, R., Carvalheiro, R., Correa, K., Figueroa, R., Lhorente, J. P., et al. (2018). Genomic prediction accuracy for resistance against *Piscirickettsia salmonis* in farmed rainbow trout. *G3* 8, 719–726. doi: 10.1534/g3.117.300499
- Yoshida, G. M., Lhorente, J. P., Carvalheiro, R., and Yáñez, J. M. (2017). Bayesian genome-wide association analysis for body weight in farmed Atlantic salmon (*Salmo salar* L.). *Anim. Genet.* 48, 698–703. doi: 10.1111/age.12621
- Yu, H., You, X., Li, J., Zhang, X., Zhang, S., Jiang, S., et al. (2018). A genome-wide association study on growth traits in orangespotted grouper (*Epinephelus coioides*) with RAD-seq genotyping. *Sci. China Life Sci.* 61, 934–946. doi: 10.1007/s11427-017-9161-4
- Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., et al. (2012). The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490, 49–54. doi: 10.1038/nature11413
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Gutierrez, Matika, Bean and Houston. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.